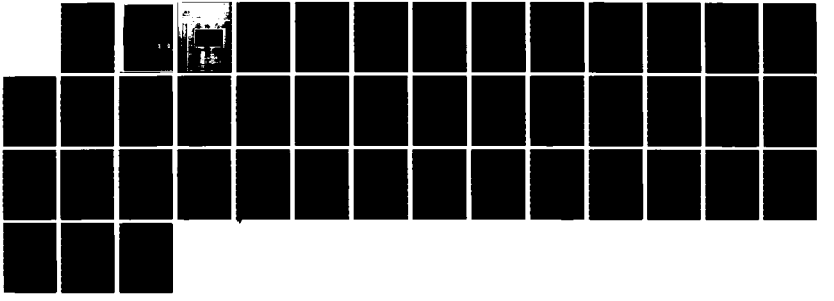
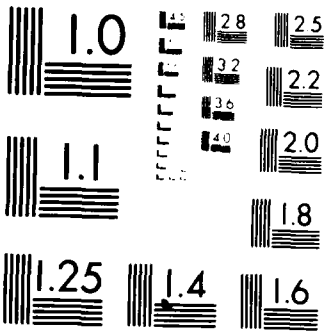
AD-A170 759 A MODEL FOR THE PROCESSING OF POSITION INFORMATION IN THE HUMAN VISUAL SY..(U) YALE UNIV NEW HAVEN CT DEPT OF OPHTHALMOLOGY AND VISUAL SCIEN.. J HIRSCH ET AL. 1/1

UNCLASSIFIED SEP 83 TR-0304 AFOSR-TR-86-0540 F/G 5/10 NL





(2)

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER AFOSR.TR. 86-0540	2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER	
4. TITLE (and Subtitle) A MODEL FOR THE PROCESSING OF POSITION INFORMATION IN THE HUMAN VISUAL SYSTEM	5. TYPE OF REPORT & PERIOD COVERED technical report
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Joy Hirsch	8. CONTRACT OR GRANT NUMBER(s) F49620-83-C-0026
9. PERFORMING ORGANIZATION NAME AND ADDRESS Yale School of Medicine, Dept. of Ophthalmology 310 Cedar Street - BML 225 New Haven, CT 06510	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 61102F 2313 A5
11. CONTROLLING OFFICE NAME AND ADDRESS Air Force Office of Scientific Research/NL Bolling AFB, DC 20332	12. REPORT DATE September 1983
	13. NUMBER OF PAGES 39
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)	15. SECURITY CLASS. (of this report)
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Spatial vision, spatial discrimination, Scaled Lattice Model, neural interpolation, retinal sampling

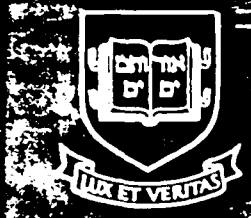
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A Scaled Lattice Model to account for the processing of positional information in the human visual system is presented. Of particular significance is the emphasis on the retinal mosaic as a two-dimensional spatial sampling system. Accordingly, the sampled image is transmitted to the cortex where it is reconstructed by neural interpolation. Two possible neural interpolation mechanisms, direct convolution and mutual

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE S/N 0102-LF-014-6601

86 8 8 080

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)



DIVISION OF IMAGING SCIENCE

A Model for the Processing of Position Information in the Human Visual System

JOY HIRSCH[1] AND RON HYLTON[2]

Technical Report Number 8304

September 1983

DIAGNOSTIC IMAGING

YALE UNIVERSITY

NEW HAVEN, CONNECTICUT

DISCLAIMER NOTICE

THIS DOCUMENT IS BEST QUALITY PRACTICABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

A Model for the Processing of Position Information in the Human Visual System

JOY HIRSCH[1] AND RON HYLTON[2]

Technical Report Number 8304

September 1983

[1] Yale University School of Medicine
Department of Ophthalmology and Visual Science
310 Cedar Street, BML 219
New Haven, CT 06510

[2] Columbia University
Department of Physics
Nevis Laboratories
New York, New York 10027

A Model for The Processing of Position Information in the Human Visual System

We present a model for the processing of positional information in the human visual system, with particular emphasis on visual tasks that involve the measurement of spatial separation. The model is in many respects a radical departure from current thinking about problems in vision. Of particular note is the fundamental significance we attach to the retinal photoreceptor lattice, considered as a two-dimensional spatial sampling system. Mechanisms of neural interpolation are discussed and hyperacuity is a natural consequence of the model. Major concerns which we do not address are questions of temporal dependence and the integration of binocular information. We refer to the model as the scaled lattice model.

The model consists of a number of elements summarized in Figure 1 as follows:

- an external luminance distribution (image) is blurred by convolution with a point spread function (or line spread function for a one dimensional distribution) and then sampled by a lattice of retinal photodetectors with center-to-center spacing d (Fig 1: a,b,c).

- the sampled image is transmitted to higher visual centers where it is reconstructed on a neural lattice by interpolating between the samples (Fig 1: d,e).

- the effective spacing of the neural lattice is some integer (N) times finer than the spacing of the photodetectors, (Fig 1: the ratio of

photodetector spacing to the neural grid is 1/3).

- position information in the reconstructed image is extracted by determining the locations of specific stimulus features on the neural lattice, in particular the luminance peaks, (Fig 1: e,f).

- the separation between two features such as peaks is determined by counting the number of neural units between the features (Fig 1: e,f,g).

- the error in determining a separation (Δs) is proportional to the positional quantization (spacing) of the neural lattice.

- the interpolation factor N (the ratio of photoreceptor spacing to neural unit spacing) varies with the separation of the peaks in such a way that the fractional error in separation ($\Delta s/s$) is roughly constant.

The effects of retinal sampling

The following discussion assumes one dimensional images for convenience. Consider the external image. It is described by the function L(x) which gives the value of the luminance at every point x. For convenience, we assume that x ranges from $-\infty$ to ∞, although the range is in fact finite. The basic problem is this: since L(x) is a continuous function of x, an exact representation of the function would require measuring the luminance at an infinite number of points. This is clearly not feasible since an eye is constructed with a finite number of photoreceptors. Thus, the question arises of how well L(x) can be represented with a finite number of measurements. A formal answer to this question is found in the sampling theorem: if L(x) contained no Fourier components with a frequency greater than f_{max}, then L(x) could be represented exactly by measuring its values at points separated in space by

a distance of $1/(2f_{max})$. Since L(x) has a finite width W, the total number of measurements is $2Wf_{max}$, a finite number. The sampling theorem also says that there is no way to represent L(x) with fewer measurements. Thus sampling L(x) at an interval of $1/(2f_{max})$ gives the most compact possible encoding of the bandwidth limited L(x). However, L(x) is not a priori limited in bandwidth. To apply the sampling theorem, we must remove all Fourier components above f_{max}. This is accomplished in the eye by convoluting the external image with the retinal (optical) line spread function, which effectively filters out the high spatial frequency components. Thus, the sampling theorem provides a very useful guide: if the line spread function removes all Fourier components above f_{max}, then sampling the band-limited function at intervals of $1/(2f_{max})$ entails no further loss of information. In principle, the continuous band-limited function can be exactly reconstructed from the samples by use of the sin(x)/x function. However, it is not clear that it is necessary or even desirable to exactly reconstruct the band-limited function everywhere. After all, the band-limited function is already a distorted version of the original luminance function.

For some visual tasks it might suffice to reconstruct the sampled image exactly only near the luminance peaks, where the function is quadratic and relatively easy to reconstruct. Indeed certain forms of distortion in the reconstruction might be useful such as the edge distortion discussed below. However, the sampling theorem gives a good estimate of what is possible, and we assume the basic reasoning is applicable to the human visual system: the line spread function acts to band limit the external luminance distribution in such a way that the image

falling on the retina can be exactly recovered from the sampled version of the image. To see more clearly what is lost and gained by this process consider Fig 1a and 1b. Fig 1a shows the external image, a narrow line (delta function). Fig 1b shows the image falling on the retina after convolution with the line spread function. The original narrow image has been spread out and now has a shape dictated by the line spread function. Thus at the initial stages of image reconstruction we have lost the ability to measure widths narrower than the line spread function, and also the ability to resolve two lines whose separation is less than the width of the line spread function. Note however that the position of the peak of the distribution in Fig 1b coincides exactly with the position of the original line, and thus no positional information has been lost, as long as the peaks of the external distribution are well separated. From this point of view, the combination of convolution with the line spread function followed by retinal sampling has effectively compressed almost all the positional information of the original distribution into a finite number of measurements for transmission to the cortex, where the ability to recover the position of the original line in Fig 1a will only be limited by the ability to locate the peak of the distribution in Fig 1b.

The above argument can be restated without reference to the sampling theorem. If the line spread function were infinitely narrow, the luminance at a sample point would contain information about the external image only at that point. With a finite width line spread function, the luminance at each sample point contains information about the external image extending over a range of positions around that point. Since the luminance at location x is now highly correlated with the luminance at $x + \Delta x$, if Δx is

much less than the width of the line spread function, there is no need to measure the luminance at $x + \Delta x$. Thus a finite sampling interval is adequate. A useful analogy is that the photoreceptors convert a continuous distribution into what is essentially a histogram. The center of a peak on a histogram can be determined quite precisely (to an accuracy much smaller than the bin size) as long as the width of the peak is several bins. (If the peak fell entirely within one bin, we could not determine its position to an accuracy better than the bin width). The line spread function acts to guarantee that the minimum width of the peak is several bins. We emphasize that the image must be blurred in some manner so that a narrow external image excites at least three photoreceptors. (The number three arises because a quadratic has three coefficients.) If not, the peak position could not be recovered from the sampled measurements. It is estimated that in the central fovea the line spread function covers approximately four photoreceptors and thus this condition is met.

Peaks and Position Information

We regard luminance peaks (maxima) as important features in the analysis of an image. There are a number of reasons to single out peaks for special consideration. One reason is illustrated in Figs 1a and 1b. As discussed above, there is a close connection between the peak in 1b and the position of the line in 1a. In fact, regions where the luminance rises or falls continuously, or remains constant, seem to contain little information about 'positions'. 'Position' generally seems to be associated with a change in the luminance slope, and a peak occurs when the sign of the slope changes. Peaks have a number of other desirable properties.

Peaks are regions where a quadratic approximation provides a good fit to the measurements, and thus peaks are quite easy to reconstruct accurately. Since a quadratic function has three coefficients, reconstructing a quadratic only requires input from three sample points. In contrast to this a function like sin(x)/x is a high order polynomial and requires many inputs. Similarly, peaks (being local features) are determined by comparing the luminances at nearby points, and a peak detector (Fig 1f) only requires input from three adjacent neural units. In contrast to this an edge or 'zero crossing' detector requires as input both the value of the luminance at each point and the value of 'zero' at each point. The 'zero' point is not a local property of the image but rather would have to be defined by averaging the image over a broad region in some manner. This is a considerable complication and it is hard to see how it could be done with sufficient accuracy. Adding a constant luminance to an image does not affect the peaks, and since adding a constant luminance gradient shifts peaks by an amount inversely proportional to the curvature of the peak, peaks of the same curvature are shifted by equal amounts, leaving their separation unchanged. Thus slowly changing luminance backgrounds have little effect on the peaks. Another interesting property of quadratics is that the sum of a set of quadratics with equal curvatures but different centers will be a single quadratic whose peak coincides with the mean of the centers of the original quadratics. This indicates that an unresolved image (one narrower than the line spread or interpolation function) will have only a single peak which coincides with the mean (center of gravity) of the luminance distribution.

There is, however, at least one problem with concentrating on luminance peaks for position information. An edge or step function certainly has a definite position, but there are no peaks to mark it. There is an interesting way around this problem. Suppose the interpolation function had a shape like that in Fig 2a, which is roughly the shape appropriate for quadratic interpolation. The rest of Fig 2 shows what happens to a step function (2b) after being convoluted with the line spread function (2c) and then reconstructed (2d) using the interpolation function of Fig 2a. The interpolation causes overshoots and undershoots, like Mach bands, to occur and thus produces a peak to mark the edge. Thus the 'imperfect' interpolation has a very useful side effect, the association of a peak with a luminance edge in the reconstructed function. Interestingly, these induced peaks do not coincide with the original stimulus edges but are systematically shifted. This introduces a systematic error in the position of an edge that might be measurable. (However, since the error is largely systematic, it can in principle be corrected for.)

The interpolation mechanism

In this section we discuss the mechanism which performs the reconstruction of the sampled image. We assume this process is linear in the sense that the output is a linear function of the inputs. Then the output can be considered as the convolution of the (sampled) input function with an interpolation function. The interpolation function gives the weight assigned to each input (photoreceptor) as a function of the distance between input position and output (neural unit) position. There are (at least) two quite different methods to perform this interpolation.

Direct convolution. The first, which we call interpolation by direct convolution, is illustrated in Fig 3a. In this scheme each neural unit (circles) receives input directly from the nearest photoreceptors (triangles). We have drawn each neural unit with three inputs, since this is the minimum number required for quadratic interpolation, and the figure shows a neural lattice with an interpolation factor (N) of 3. The inputs are summed with appropriate weights to implement the interpolation function. We can see that this scheme is characterized by a large number of connections per photoreceptors (3N) and a large number of different weights (3N/2 rounded upwards).

Mutual coupling. An alternative scheme is presented in Fig 3b. We call this interpolation by mutual coupling, and it is similar to existing models of lateral inhibition (Ratliff and Hartline, 1959μ Ratliff, Hartline and Miller, 1963). Again each neural unit typically receives input from three sources: itself and its two nearest neighbors. Each photoreceptor also provides input for the neural unit immediately below it, but not for any other. The neural units form a mutually coupled system obeying a second order difference equation, with the inputs from the photoreceptors providing the excitation function. The solution of such a system is given by the convolution of the excitation function with the impulse response of the mutually coupled system, which is exactly the desired result.

This scheme has a number of significant advantages over the first. There is only one connection per photoreceptor independent of the interpolation factor N. The number of different weights is also independent of N. There is one weight for the photoreceptor connection, a

second for the connection between a neural unit and itself, and a third for the connection between a neural unit and its nearest neighbors. Another weight would only be required if each neural unit had to be connected to its second-nearest neighbors, which would result in a fourth order difference equation. Another important advantage in the mutual coupling scheme is that all neural units are the same, except for the additional photoreceptor input which some receive. In the direct convolution scheme, the set of weights used by each neural unit depended on the position of the neural unit relative to the photoreceptors, which requires that each neural unit 'know' its position to interpolate properly. This complication does not arise in the second scheme where all neural units are essentially the same. That is, for the direct convolution, the interpolation function was explicitly contained in a set of weights that depended on position, while in the mutual coupling case the interpolation is done implicitly by the impulse response of the second order system. There is one further advantage to the second scheme. Suppose that something caused the output of one neural unit to fluctuate upwards. Since this output is an input for the adjacent neural units and we estimate that the relative weight given to nearest neighbors is positive (and approaches unity as the interpolation factor increases), these neural units will also fluctuate upwards, with the net result that the fluctuation in the <u>difference</u> between adjacent neural units will be highly suppressed. Since peaks are determined only by these differences, this results in a considerable degree of noise suppression for the peak detectors, a suppression that does not occur in the first scheme. A natural consequence of the mutual coupling scheme is that the interpolation factor (the ratio of photoreceptor to neural unit spacing)

must be an integer. There cannot be a fractional number of neural units between two photoreceptors. The direct convolution scheme only requires that the interpolation factor be the ratio of two integers to guarantee that the neural lattice is coherent with the photoreceptor lattice.

However, the mutual coupling system as illustrated in Figure 3b is oversimplified since a one dimensional system with only nearest-neighbor couplings cannot have an impulse response as shown in Figure 2a. However, a two dimensional system on a hexagonal lattice probably could have such an impulse response, at least for certain orientations, and a hexagonal system with nearest-neighbor and second-nearest-neighbor couplings almost certainly could. Since the primate photoreceptor array has been shown to be a highly regular hexagonally packed lattice (Miller, 1979μ Hirsch and Hylton, 1983) we consider the mutual coupling system to be a viable notion.

Now let us consider the constraints that can be placed on the interpolation function. The most important is that peaks should be reconstructed accurately, and this will be true of most any symmetric function of reasonable width which is itself quadratic near $x = 0$. We estimate this width as follows. As noted above, quadratic interpolation in one dimension requires input from at least three photoreceptors. This sets a minimum 'range' of about +1.5 photoreceptors. A two dimensional quadratic has six coefficients, and thus requires at least six inputs. If each photoreceptor has six neighbors, then a 'range' of 1.5 photoreceptors is again sufficient. Suppose the range were greater than this. If the image were indeed quadratic, this would allow a more accurate constrained interpolation. However, if the image were not quadratic, this would result

in the peak at a particular position being shifted around by the luminance at a relatively distant point. If we assume that two fully resolved lines should not interfere with each other, this sets a maximum range roughly comparable to the range of the line spread function for very fine separations. Then, the interpolation functions might look like Fig 2a with a total width comparable to the line spread function, say about 1 or 2 minutes from the positive peak to the negative peak. However, this assumption does not seem to be correct (i.e., there are 'crowding effects').

A series of experiments reported by Westheimer et al.(1975) may provide a better estimate of the width of the interpolation function. Westheimer has shown that the ability of observers to discriminate the vernier offset of a pair of vertical lines or the orientation of a single vertical line is susceptible to interference from flanking lines, and that the interference is maximum when the flanking lines were offset from the target lines by about 2 - 3 minutes of arc. If we attribute the interference to a conflict between the negative sidelobes of the flanking lines and the positive central lobe of the target lines (perhaps by causing multiple peaks or some other non-quadratic effect), this indicates that the negative lobes are about 2 -3 minutes from the central positive lobe, reasonably comparable to the previous guess. Indeed, if the exact mechanism of the interference were understood, this experimental technique might provide a fairly direct method for determining the exact shape of the interpolation function.

As discussed below, the interpolation factor N varies in such a way that the spacing of neural units and the jnd in separation are roughly

proportional to the separation between luminance peaks. This raises the possibility that the width of the interpolation function is not constant but rather scales with peak separation (or rather with N). This would naturally follow if peaks and edges tended to become wider as the separation between them increased.

Separation measuring and scaling

In this section we discuss the process of measuring separations between peaks. This discussion is largely based on results of an experiment which effectively measured the fractional jnd in separation ($\Delta s/s$) as a function of the inverse separation ($1/s$) between two peaks (Hirsch and Hylton, 1982). (The jnd for discriminating between spatial frequencies was also measured as a function of spatial frequency, and those results were also found to depend only on the distance between two peaks in the spatial frequency grating. An example of the results appears in Fig. 4. The relation between frequency and separation is $f = 1/s$ and $\Delta f/f = \Delta s/s$). It was found that Δs tended to take on discrete values given by $1/N$ times the photoreceptor spacing, and that the value of N varied with s in such a way that $\Delta s/s$ was roughly constant. Specifically, it was found that $\Delta s/s$ as a function of $1/s$ consisted of segments of constant Δs with Δs given by $.008°/N$ within a segment, and that the transition between different segments occurred at intervals that were equally spaced in $1/s$ and coincided with a maximum value of $\Delta s/s$ of approximately .032. On the average, $\Delta s/s$ had a value of .025.

The first step in separation measurement is to locate the positions of the peaks. This is illustrated in Fig 1f where each neural unit, plus the

two adjacent neural units, is fed into a peak detector (squares). The peak detector compares the two successive differences between the three neural units and decides if the sign of the slope has changed. The output of the peak detector is a logical signal indicating that a peak has occurred within +1/2 neural unit of the current position. Thus the effective spacing of the neural units introduces a fundamental limit to the accuracy with which a peak can be located, the positional quantization error of the neural unit lattice. The location of the second peak is similarly determined to an accuracy of +1/2 neural unit. The distance between the two peaks is then determined to an accuracy of +1 neural unit. This reflects the essentially digital nature of the position information: separation is measured by counting the number of neural units separating the two peaks, and the counting process is subject to a +1 count uncertainty. Alternatively, the quantization error can be attributed to the 'rounding' of the peak position to the nearest neural unit position.

It is clear that the limit of resolution in separation is proportional to the spacing of the neural units. If the neural unit spacings take on the values d/N as described previously, then the jnd in separation will be proportional to d/N. Experimentally, we find that $\Delta s = .008°/N$, where Δs is defined as the difference in separation necessary to correctly discriminate between two different separations 75% of the time. If we assume the psychometric function is gaussian, then a 75% correct response corresponds to a mean displacement of .68 standard deviations. Thus the rms error is $.008°/(.68) = 1.47 * .008°$. Anatomically, the photoreceptor spacing d is estimated to be about 30 seconds or .008° Westheimer, 1979), and hence the rms error is 1.47 * d or 1.47 counts. Since the maximum

quantization error is only +1 count, an rms value of 1.47 is clearly larger than expected suggesting that there may be other sources of error which are proportional to the neural unit spacing. (We would expect the rms error to be about .5-.7 counts for pure quantization error). An alternative explanation might be that the peak detectors are spaced every two neural units rather than every neural unit, doubling the quantization error.

Scaling

Results of the cited experiment show that the jnd in separation Δs is not constant but rather varies with the inverse separation (1/s) being discriminated in such a way that $\Delta s/s$ is roughly constant. That is, the value of N is automatically chosen to keep the neural unit spacing roughly proportional to the separation between the peaks, which is generally described by saying that Δs scales with s.

Suppose two reconstructed peaks are separated by a distance s on a neural lattice with spacing Δx. Then the separation in terms of neural units (counts) is given by $L=s/\Delta x$ with a +1 neural unit uncertainty. The fractional uncertainty in separation is given by $1/L = \Delta x/s$. From this it can be seen that requiring the fractional error to be constant is essentially the same as requiring that the number of neural units spanning the two peaks (L) be constant. As s becomes small the fractional error becomes large, and at some point reaches a maximum tolerable value of $c_{max} = 1/L_{min}$. (Numerically, c_{max} is found to be .032 and hence $L_{min} = 30$.) To keep performance acceptable for smaller values of s the value of Δx must be decreased. As discussed above, the nature of the

interpolation mechanism is such that Δx has the form d/N where d is the photoreceptor spacing and N is the integer interpolation factor. Then the fractional error is given by (d/N)/s and the minimum value of N for a given s is $N = L_{min}\ d/s$, rounded upwards.

One question that might arise at this point is why N varies at all. At first glance it might seem simpler to keep N at a large fixed value. The most direct answer to this is that experiments show that N does vary and in fact takes on the value required to keep $\Delta s/s$ constant. There are two arguments to justify this behavior. The first comes from a consideration of the errors introduced in separation measurement by errors in the spacing of photoreceptors. Assume that the photoreceptor lattice is constructed by placing end to end intervals of mean width d and error σ. The different intervals are assumed to be independent of each other. Then after we have placed down L intervals we have covered a mean distance of $s = Ld$ with an accumulated error of σ*SQRT(L) = σ*SQRT(s/d). Thus the accumulated separation error on the photoreceptor lattice increases proportionally to SQRT(s), and this error will exceed any fixed error as s increases. The essential point is that there exist sources of error which increase as s increases and thus Δs must also increase with s. If we assume that objects have some natural width which increases as the separation between objects increases (i.e., that the boundaries of objects become bigger as the objects become bigger) then this gives another source of error that increases with s.

The photoreceptor spacing argument can be inverted to place a limit on the fractional rms error in photoreceptor spacing (σ/d). We find experimentally that the average jnd in separation is given by $\Delta s = \alpha * s$

where $\alpha = .025$. As discussed above, the jnd corresponds to .68 times the rms error in s. Thus if the photoreceptor spacing error is not to dominate Δs, we require that the contribution of the photoreceptor error to Δs is less than 1/2 the observed value of Δs or

$$.68 * \sigma * \mathrm{SQRT}(s/d) < 1/2 * \alpha * s \text{ or}$$

$$\sigma/d < 1/2 * 1/.68 * \alpha * \mathrm{SQRT}(s/d).$$

The average value of $L = s/d$ is $1/\alpha = 40$ requiring that $\sigma/d < .12$, a reasonable number. Assuming a minimum separation of 4 photoreceptors ($s/d = 4$) requires $\sigma/d < .037$, a fairly small number. Perhaps there exists a small region of the fovea in which photoreceptor spacing error is small, and the measurement of small separations could be restricted to this region.

As an alternative to requiring accurately spaced photoreceptors, it is conceivable that the system includes a 'self calibration' feature that can correct for errors in photoreceptor spacing. This is possible since the spacing errors cause systematic, position dependent, errors in s. If s is measured through the AND gate mechanism illustrated in Fig 1g, the calibrations could be done by sweeping a pair of peaks of fixed separation across the retina and marking all AND gates which respond as belonging to the same value of s. This limits the error contribution from photoreceptor spacing to about one neural unit since a spacing error greater than this will result in the AND gate being classified as responding to a different value of s. Note that this error is still proportional to the neural unit spacing. (Similarly, since any errors in the interpolation weights represent systematic errors, it is possible to correct them.) However, while such corrections are conceivable, we feel that it is more likely that

the photoreceptor lattice is constructed with sufficient accuracy to begin with.

The second argument in favor of scaling is that having a number of different N's does not necessarily increase the complexity of the whole system, defined as the number of parts required to build it. In fact, if one attempts to count the total number of neural elements required to construct a hardware separation detector, it turns out that a system with $\Delta s/s$ constant actually requires vastly fewer elements than one with Δs constant. This occurs because the system must check all pairs of peak detectors and the system with scaling has many fewer pairs. This is illustrated in Fig 1g. The separation detector must perform the operation 'Peak(x) AND Peak(x + s)' OR'ed over all starting positions x. To see the effect of scaling, assume that the left hand peak position is fixed at x = 0. Then the number of pairs (AND gates) required to cover all possible position for the right hand peak from s_{min} to s_{max} is given by

$$\int_{s_{min}}^{s_{max}} \frac{ds}{\Delta s}$$

If we take Δs as the constant $\Delta s_{min} = \alpha * s_{min}$ then the number of pairs is $(1/\alpha)(s_{max}/s_{min})$. On the other hand, if Δs varies with s as $\Delta s = \alpha * s$, the integral becomes

$$(1/\alpha) \ln (s_{max}/s_{min}).$$

For large s_{max}/s_{min} the system with scaling requires many fewer pairs than the one without.

The above analysis corresponds to a one dimensional system in which the image is required to be exactly centered in the visual field, and is

only meant as an illustration. The correct analysis requires knowledge of how the neural units are distributed in two dimensions as a function of eccentricity. Actually, the above case is relatively unfavorable to the scaling mechanism. The true advantage of the scaling mechanism is probably much better than $x/\ln(x)$. Extending the above argument to two dimensions introduces an extra factor of $s/\Delta s$ under the integral sign. Then with no scaling the number of elements is $1/2\ (1/\alpha)^2\ (s_{max}/s_{min})^2$ but with scaling the number is $(1/\alpha)^2\ \ln\ (s_{max}/s_{min})$, still only a logarithmic increase.

It should be noted that the notion of scaling presented here differs fundamentally from the usual notion of scaling with eccentricity. The more conventional view is that there exists a single non-linear mapping between the retina and the cortex, with a step in distance of constant size on the cortex corresponding to a step in distance on the retina whose size increases with eccentricity. In contrast to this, we suggest that there exists a set of linear mappings from the retina to the cortex with different scales or magnification factors. While the different mappings have different magnification factors, within one map the magnification factor is constant. The physiologically and psychophysically observed increase in magnification factor with eccentricity presumably occurs because progressively finer mappings are restricted to progressively smaller eccentricities. This corresponds to requiring that images be roughly centered in the visual field. Then the boundaries of wider objects tend to fall at larger eccentricities while small objects lie within small eccentricities, and the fine maps are only needed at small eccentricities. Further, if we assume that roughly equal neural resources are allocated to

each map, then the total area that can be covered by a map decreases as its fineness increases. As a function of eccentricity, we might expect As to be constant over some central region covered by the finest map, and then increase in a series of steps as the boundaries of successively coarser maps are exceeded, with As roughly proportional to eccentricity. Note that there is no a priori reason for the various maps to be circularly symmetric. The scaled linear maps have tremendous advantages over a nonlinear map in terms of the simplicity of image analysis and pattern recognition. It is also difficult to accurately measure distances on a nonlinear map since this requires integrating variable magnification factors along some path which is likely to be difficult to define.

Scaling at Low Frequencies

The discussion of scaling above was basically addressed to a high frequency or hyperacuity region involving neural lattices with spacing equal to or less than the photoreceptor spacing. Given the advantages of a scaled system over a non-scaled system, it would seem quite reasonable to assume that the scaling property is general and applies to all image sizes or spatial frequencies accessible to the human visual system, say angular sizes of 90° to 1/30°, and preliminary measurements indicate that this is indeed the case. We now discuss how scaling might be accomplished at low frequencies. At high spatial frequencies the effective neural spacing is d/N where d is the photoreceptor spacing and N decreases with decreasing frequency in order to keep d/N proportional to s = 1/f. This scheme will fail eventually simply because there are no positive integers less than 1, and indeed a breakdown in the d/N dependence is observed to occur below 2

cycles/deg, with Δs remaining roughly proportional to s rather than becoming a constant equal to d. In fact, experimentally we observe that for foveal measurements between about 1 and 2 cycles/deg, Δs takes on the value $d * 2$ which we interpret as evidence for a neural lattice with spacing twice the photoreceptor spacing. Thus at low frequencies the function of the scaling system is to reduce an excessively fine retinal photoreceptor lattice to a coarser neural lattice which is presumably more tractable since it has fewer elements. There are two straightforward mechanisms for accomplishing this, one possible only in the central region of the retina (defined as the region where each photoreceptor has its own connection to the cortex) and the other suitable for foveal and also peripheral regions where there are many fewer cortical connections than photoreceptors.

In the central region where there exists one connection per photoreceptor the basic task is simply to produce a coarser lattice. This could be done simply by constructing a neural lattice whose spacing is N times coarser (not finer) than the photoreceptor spacing and using an interpolation mechanism like Fig. 3b in reverse, with 1 neural unit output per N photoreceptor inputs rather than N outputs per each photoreceptor input, with a broad interpolation function to provide pooling over N photoreceptors rather than an interpolation between photoreceptors. In this case Δs takes on the form $d * N$ as opposed to the previous case where Δs was d/N, with the value of N increasing to keep $\Delta s/s$ constant as s increases. Note that N now increases with s to keep $\Delta s/s$ from getting too small, whereas previously N increased as $1/s$ increased to keep $\Delta s/s$ from getting too large.

The above scheme is only feasible in the central region where there already must exist one cortical connection per photoreceptor to allow 1/N interpolation. In peripheral regions where the hyperacuity interpolation mechanism does not exist it would be foolish to transmit the output of every photoreceptor to the cortex only to discard most of the information. A much more elegant scheme would be to pool groups of photoreceptors into spatial sampling units, effectively forming large photoreceptors, and transmit only one output per each pool to the cortex. As long as the widths of the pools and the center-to-center spacing of the sampling points (pool centers) is such that a narrow feature contributes to at least three sample points one can then invoke a cortical interpolation mechanism identical to the one that interpolated between individual photoreceptos to now interpolate between the sampling pools. Indeed, the whole mechanism is identical to the one discussed for high frequencies except that the line spreading is now accomplished by retinal pooling rather than optical blurring, and the retinal lattice is formed by the centers of the pools rather than by individual photoreceptors. The retinal pooling mechanisms might look like Fig. 3b in reverse, as above. If the center-to-center spacing of the pools is K times the photoreceptor spacing d, then Δs has the form Kd/N where K is the retinal pooling factor and N is the cortical interpolation factor.

Preliminary results suggest that the second scheme is indeed employed at low frequencies, at least for frequencies between .25 and 2 cycles/deg (image sizes between .5° and 4°) in the fovea. Within this range Δs appears to have the form Kd/N where d is .008°, K is 8, and N takes on the values 1 through 4. (That is, there are 4 segments to the $\Delta f/f$ curve

between .25 and 2c/deg.) This is slightly different from the high frequency region where N took on the values 1 through 7 or 8. A retinal pooling factor of 8 corresponds to a center-to-center spacing for the retinal pools of 8 * .008° = .064° or 1/16°. If we assume that there are a sequence of pool sizes each 8 times larger than the last, then image sizes between 32° and 4° will be handled by pools with center-to-center spacing 8 * 1/16° = .5° and image sizes larger than 32° will be handled by pools with center-to-center spacing of 8 * .5° = 4°. While these numbers are somewhat speculative, it is clear that the high frequency band (K = 1, 8 segments, 2 to 32 c/deg) plus three low and intermediate frequency bands (K = 8, 64, and 512 with 4 segments each, minimum frequencies 2/K cycles/deg) provide a set of 20 neural maps spanning image sizes from 1/32° to over 180° with $\Delta s/s$ or $\Delta f/f$ never varying by more than a 2/1 ratio. The different values of K correspond to sampling by individual photoreceptors (K = 1) and to sampling by pools of photoreceptors with center-to-center spacing of K * .008° = 1/16°, .5°, and 4° respectively.

Since large images tend to be associated with high eccentricities we expect that there should be a correlation between the distribution of the sizes of the retinal sampling pools and eccentricity. In the fovea all 4 values of K might coexist with the smaller values of K disappearing as eccentricity increases and only the largest pools existing at the highest eccentricities. It will be interesting to see if this simple scheme can be reconciled with the anatomy of the retina.

Application to Other Experiments

The discussion above was specifically based on a particular experiment, the discrimination of the separation between a pair of parallel lines as a function of the separation. Our model must of course be very general if it is to be useful. We present here an application of the model to some other experiments, albeit also very simple ones. Application of the model to more complicated visual tasks probably would not require any fundamental modification.

Separation-discrimination tasks can be regarded as having two independent characterizations. The first is the spatial scale s and the second is the jnd or resolution scale Δs. For the parallel line task s was taken to be the perpendicular distance between the two lines and Δs was the jnd in separation also measured along the perpendicular. The parallel line task is a special case in which the perpendicular separation between the lines is both the stimulus variable used to measure Δs and also the scale-setting distance.

A quite different situation prevails in the classical vernier discrimination task which measures the jnd for aligning two pointers separated by some gap. In this case the spatial scale s is defined by the length of the gap between the two pointers whereas Δs is the jnd lateral offset between the pointers and is measured perpendicular to the gap. Thus s for the vernier discrimination task is not the stimulus variable but rather a parameter of the experiment. Further, s and Δs are measured along orthogonal directions. Thus, outside the context of our model, it is by no means obvious that there should be a direct connection between s and Δs (gap length and vernier threshold). Nevertheless, we predict that Δs for

vernier discrimination should be the same as Δs for line separation discrimination if the gap length equals the parallel line separation, neglecting orientation effects discussed below. Preliminary results confirm this prediction. The connection between the two experiments can be seen more clearly if we reduce the lines in the two experiments to small dots. Then the parallel line task measures Δs along the line connecting the dots whereas vernier discrimination measures Δs perpendicular to the line connecting the dots. The scale s in both cases is set by the distance between the dots, and it is s which determines Δs for both directions (neglecting orientation effects). This example indicates that applying our model to other experiments mainly requires the identification of the scale-setting distance s, which may not be easy for complicated tasks.

There is another concern in applying the model to small separations. Much of the discussion above was based on an experiment 'which effectively measured the ability of observers to discriminate the separation between two parallel lines'. We included the word 'effectively' because that experiment mainly measured the ability of observers to discriminate between sinusoidal spatial frequency gratings of different spatial frequencies. We have shown that spatial frequency discrimination is the same as discriminating the separation between two lines with separation $s = 1/f$, where f is the spatial frequency. However, sinusoidal gratings possess one sterling virtue that makes them more suitable for studying small values of s (high values of f) than the corresponding line pair (or dot pair) would be. This virtue is that no linear operation (line spread, interpolation) can change the shape of the grating, but rather can only change the contrast and shift the phase. Thus the peaks of a grating of frequency f

are always separated by s = 1/f, no matter how small s gets (when it gets too small, the grating contrast becomes too low for the grating to be seen). In contrast to this, as two lines get close together, their profiles (as determined by the line spread and interpolation function) begin to interfere so that the separation between the two peaks will no longer simply be the separation between the external narrow lines. When they get too close together there will only be one peak and the two lines will appear to be a single fat line. Thus in practice there will often be critical minimum separations and line lengths in an experiment (on the order of a few minutes) and our model might break down if these conditions are not met. We also find that failing to randomize the positions of stimuli can produce unexpected results, and that a repetitive stimulus such as a grating generally gives better defined segmentation than a non-repetitive pattern such as a single pair of lines, possibly because the repetition provides stronger excitation to the scale setting mechanism.

Orientation Effects

There is one further point that needs to be made in our model which does not appear in any of the one-dimensional arguments given previously. The problem is the identification of the perpendicular distance between two parallel lines as the appropriate definition of their separation. While this is a simple and appealing definition it is by no means obviously correct. One could certainly imagine that there existed certain intrinsic measurement directions fixed with respect to an observer's head orientation and that separations are always measured along one of these natural directions rather than along a direction defined by the stimulus itself.

This leads to the possibility that both s and As are dependent on the orientation of the stimulus relative to the observer. Since we have assigned the photoreceptor lattice a fundamental role in our model, it would seem necessary to assume that the two dimensional characteristics of the photoreceptor lattice must be of importance. The fundamental two-dimensional characteristic of photoreceptors is their hexagonal packing and the likely consequence of this for our model is that there exist intrinsic directions for measuring separation and that these directions are separated in orientation by 60°. This suggests that the effective distance between two parallel lines is the perpendicular distance divided by the cosine of the angle between the perpendicular direction and the closest intrinsic direction. Since this angle cannot be greater than 30°, the maximum possible effect is $1/\cos(30°) = 1.15$. Thus, the effective distances may be larger than the perpendicular distances by as much as 15% (or spatial frequencies lowered by 15%) as orientation is varied. Since the effect is systematic, it can in principle be corrected for and thus may not be measureable. However, if changing the orientation caused the scale setting separation to pass through a transition from one value of N to another, the resulting change in As would be measureable, and indeed we have observed this effect (Hirsch and Hylton, 1983).

Stereo Vision

We have not considered stereo vision in detail. However, much of the model seems directly transferable to stereo acuity tasks. We make this argument based on two results reported by Butler and Westheimer (1978). First, stereo disparity thresholds are susceptible to interference from

flanking lines in direct analogy to the interference of flanking lines on vernier acuity thresholds and line orientation thresholds. Since we attribute this interference to the interpolation mechanism, we take this as evidence that sampling and interpolation as discussed above applies to stereo vision.

Butler and Westheimer (1978) also report that stereo disparity thresholds increase as disparity increases, which he describes as a disparity tuning function. We interpret this as the analog of $\Delta s/s$ scaling for stereo vision where Δs is identified as the jnd in disparity and s is disparity. We suggest that disparity threshold as a function of disparity is analogous to line separation threshold as a function of line separation and vernier offset threshold as a function of gap length. The rise of disparity threshold with disparity is not the result of a disparity tuning mechanism but rather a consequence of the ubiquitous scaling mechanism, with disparity identified as the scale setting variable for stereo acuity tasks. This might be envisioned as two parallel lattice planes, one for each eye, stacked vertically and registered in such a way that features with zero disparity are vertically aligned. Disparity is then the distance measured parallel to the lattices between a point in one lattice and the corresponding point in the other, whereas a spatial separation is measured between two points in the same plane, establishing the analogy between disparity and spatial separation. Then disparity thresholds should scale with disparity just as spatial thresholds scale with spatial variables.

Summary

We have presented a model for the processing of positional information

in the human visual system. The model can be summarized as follows.

An external luminance distribution or image is blurred by convolution with the line spread function so that a narrow line is spread over at least three photoreceptors. The blurred image on the retina is then sampled by the photoreceptor lattice and transmitted to higher visual centers. The neural represenation of the image is constructed by interpolating the sampled function onto a lattice of neural units. The reconstruction process is implemented by convolution with the interpolation function. The effective spacing of elements in the neural lattice is finer than the spacing of photoreceptors by an integer factor N. The neural lattice spacing determines the level of accuracy to which positions of stimulus features can be determined. Higher values of N correspond to higher levels of spatial resolution.

Position information in the recontructed image is extracted by determining the locations of specific stimulus features on the neural lattice, in particular the luminance peaks. The separation (s) between features such as peaks is determined by counting the number of neural units between the features. Hence the error in determining a separation (Δs) is proportional to the positional quantization error (spacing) of the neural unit lattice.

The interpolation factor N (the ratio of photoreceptor spacing to effective neural unit spacing) is chosen to keep the fractional error ($\Delta s/s$) approximately constant. More generally, the features whose positions or separations are being analyzed define a spatial scale (s) which is also used to set the scale for the spatial resolution (Δs).

Acknowledgements

This work was partially supported by grants from NEI EY00785 and EY00167, Research to Prevent Blindness, The Connecticut Lions Eye Research Foundation Association, and the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under grant number 49620-83-C-0026. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation thereon.

There is an interesting way to summarize the sampling and reconstruction system. As discussed above, the 'coarse' sampling by the photoreceptors is the most efficient encoding of the image falling on the retina. This is certainly an advantage when transmitting the information from the retina to the cortex. However, in the cortex the most efficient possible encoding is not necessarily a convenient representation of the image. The interpolation system provides a mechanism for unpacking the encoded information into a more easily analyzed form. The neural lattices of different N represent different degrees of unpacking, and the degree of unpacking selected is chosen to keep $\Delta s/s$ constant. The neural representations of different N contain no information that was not present in the original sampling. They contain the same information in a more accessible form.

Figure Captions

FIG. 1 A schematic of the model is given in Fig 1. Fig 1a shows the external luminance distribution which we have taken to be a delta function. Fig 1b shows the external image after convolution with the line spread function. Fig 1c illustrates sampling of the retinal image by photoreceptors. Fig 1d represents the interpolation process, which accepts the coarsely sampled input from the photoreceptors and produces a more finely spaced sampling on the neural unit lattice, Fig 1e. We have drawn the neural unit spacing as 1/3 the photoreceptor spacing: in general it can be 1/N. Fig 1f illustrates peak detection on the neural unit lattice. A peak detector (square) requires input from 3 adjacent neural units. Fig 1g illustrates a separation detector: Peak(x) AND Peak (x+s) is OR'ed over all x.

FIG. 2 (a) a prototype interpolation function, (b) a step function that is blurred by the line spread function, (c) and then reconstructed (d) using the interpolation function in (a).

FIG. 3 (a) Interpolation by direct convolution. (b) Interpolation by mutual coupling.
See text for explanation.

FIG. 4 Reprinted from Hirsch and Hylton, 1982 (Fig. 1). Fractional jnd in spatial frequency, $\Delta f/f$, as a function of reference frequency, f, for three observers: JH, BA and MM. The straight lines passing through the origin represent regions of constant angular jnd, Δs.

REFERENCES

Barlow, H.B. (1979) Reconstructin the visual image in space and time. Nature 279, 189-190.

Barlow, H.B. (1981) The absolute efficiency of perceptual decisions. Proc. R. Soc. Lond. 212, 1-34.

Blakemore, C. (1970) Range and depth of depth discrimination. J. Physiol., 211, 599-622.

Bracewell, R.M. (1956) Strip integration in radio astronomy. Aust. J. Phys. 9, 298-217.

Butler, T.W. and Westheimer, G. (1978) Interference with stereo-scopic acuity: Spatial, temporal and disparity tuning. Vision Res., 18, 1387-1392.

Hirsch, J. and Hylton, R. (1982) Quality of the primate photoreceptor lattice. Vision Res., in press.

Hirsch, J. and Hylton, R. (1982) Limits of spatial frequency discrimination as evidence of neural interpolation. J. Opt. Soc. Am., 72, 1367-1374.

Hirsch, J. and Hylton, R. (1983) Orientation dependence of hyperacuity contains components with hexagonal symmetry. Suppl. to Invest. Ophthal. and Visual Sci., 24(3) 276.

Miller, W.H. (1979) Reconstructing the visual image in space and time. Nature 279, 189-190.

Ratliff, F. and Hartline, H.K. (1959) The responses of limulus optic nerve fibers to pattern of illuminations on the receptor mosaic. J. Gen. Physiol., 42, 1241-1255.

Ratliff, F, Hartline, H.K., and Miller, W.H. (1963) Spatial and temporal aspects of retinal inhibitory interaction. J. Opt. Soc. Am., 53, 110-120.

Sakitt, B. and Barton, H.B. (1982) A model for the economical encoding of the visual image in cerebral cortex. Biol. Cybern 43, 97-108.

Snyder, Allan W. (1982) Hyperacuity and Interpolation by the visual pathways. Vision Research, 22, 1219-1220.

Westheimer, G. (1979) The spatial sense of the eye. Invest. Ophthal. Visual Sci. 28, 893-912.

Westheimer, G. and Hauske, G. (2975) Temporal and spatial interference with vernier acuity. Vision Res. 15, 1137-1141.

SCHEMATIC MODEL OF POSITION PROCESSING

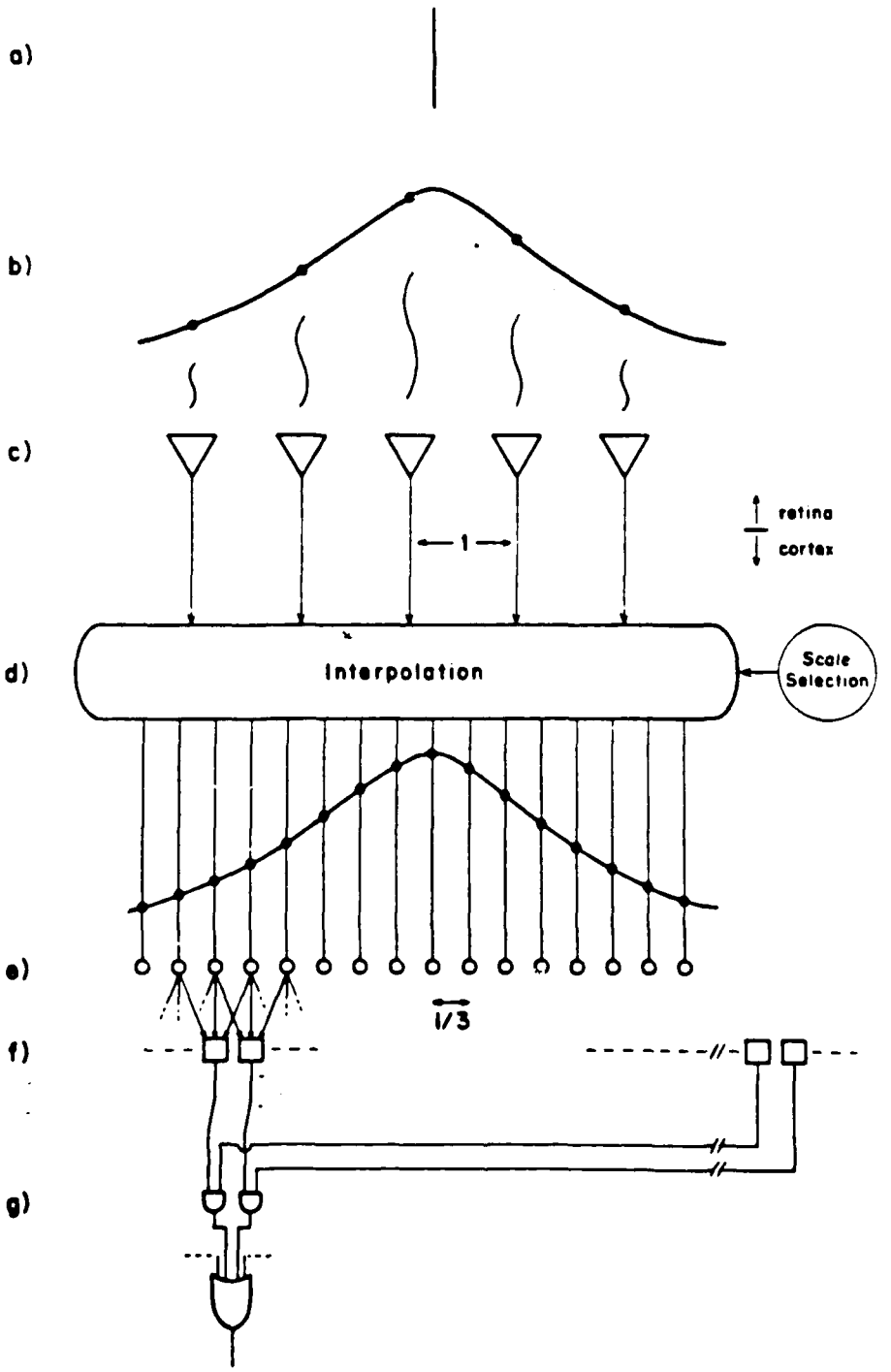


FIGURE 1

ILLUSTRATION OF EDGE RECONSTRUCTION

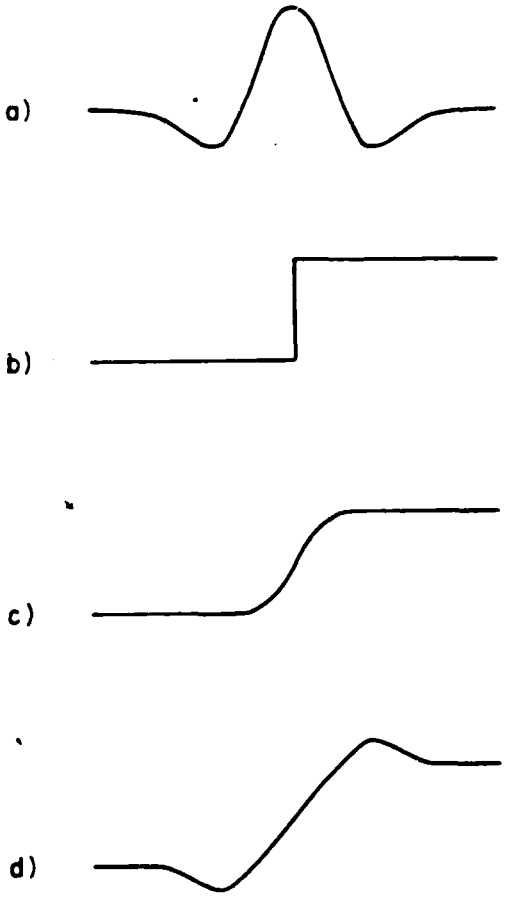
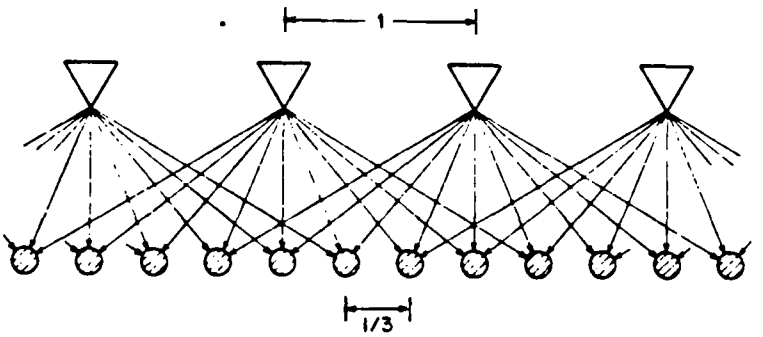


FIGURE 2

TWO POSSIBLE MECHANISMS OF NEURAL INTERPOLATION

a) Direct Convolution



b) Mutual Coupling

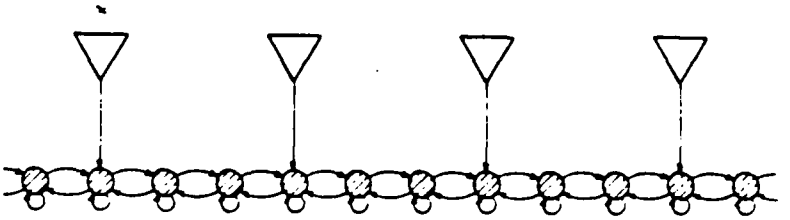
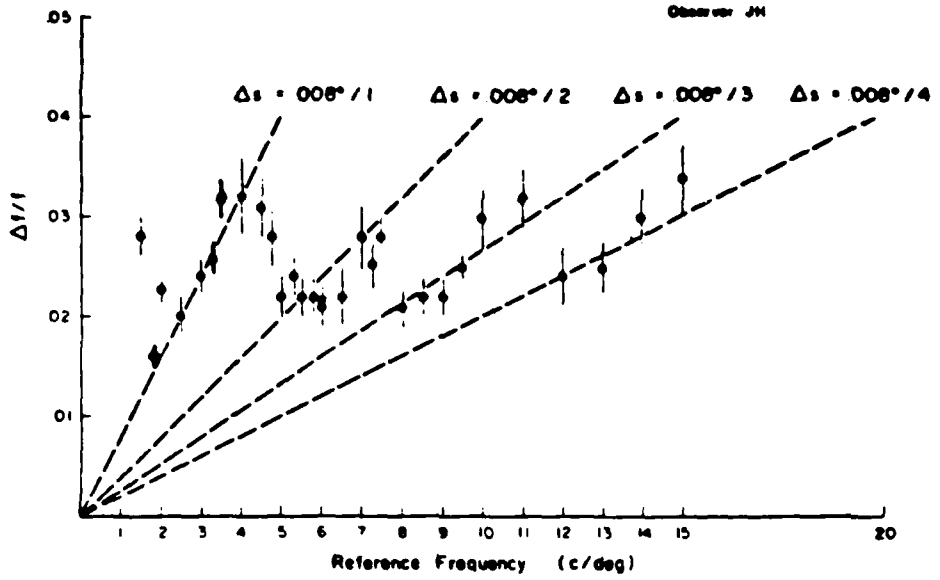
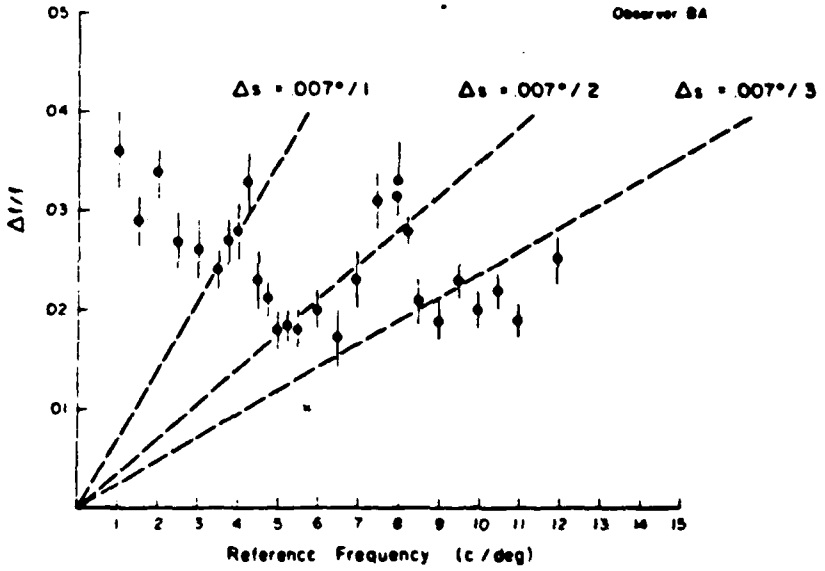


FIGURE 3

SPATIAL FREQUENCY DISCRIMINATION





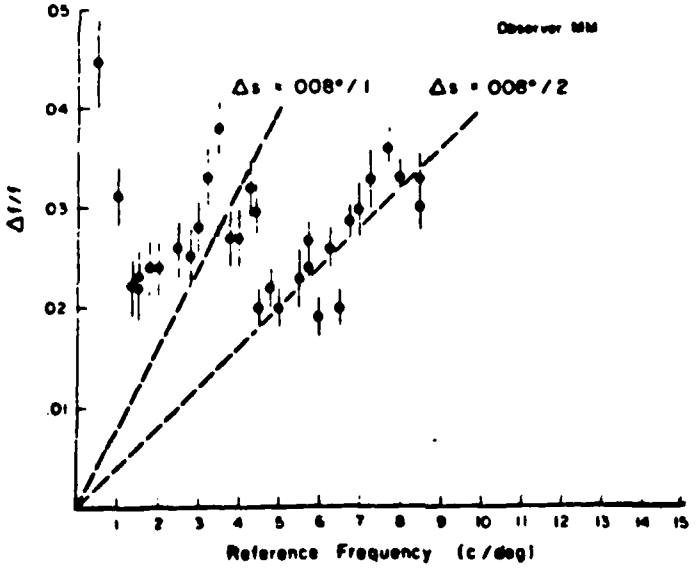


FIGURE 4

END

DTIC

9-86